©1999 The Institute of Mind and Behavior, Inc.
The Journal of Mind and Behavior
Autumn 1999, Volume 20, Number 4
Pages 403–420
ISSN 0271–0137

Internal Representations — A Prelude for Neurosemantics

Olaf Breidbach

Friedrich Schiller University

Following the concept of internal representations, signal processing in a neuronal system has to be evaluated exclusively on the basis of internal system characteristics. Thus, this approach omits the external observer as a control function for sensory integration. Instead, the configuration of the system and its computational performance are the effects of endogeneous factors. Such self-referential operation is due to a strictly local computation in a network. Thereby, computations follow a set of rules that constitutes the emergent behaviour of the system. Because these rules can be demonstrated to correspond to a "logic" intrinsic to the system, it can be shown that the concept of internal representation provides the basis for neurosemantics.

What are the basic structural properties of a nervous system? The neuroanatomist Gaze put it this way: "nerve pathways always run from here to there" (1970, p. 1). The nervous system is no statistical amalgam of integrative devices but is a topologically ordered system with local characteristics (Braitenberg and Schüz, 1991). Thus, it seems reasonable to start from this view in order to reconstruct the integrative actions of the nervous system. Such an approach, accordingly, has to work with strictly local computations. Consequently, information transfer in such a system should not be described using the idea of a representation based on an external evaluation of information transfer (Shannon and Weaver, 1949) but should follow the approach of an internal representation (Rusch, Schmidt, and Breidbach, 1996).

According to the concept of internal representation, information is to be evaluated on the basis of system-intrinsic variables. Thereby, a subjective probability is characterized which defines the effect of an input signal not with regard to the transformation of an objective probability, describing the

I thank Paul Ziche for his critical comments on the manuscript. Requests for reprints should be sent to Prof. Dr. Olaf Breidbach, Ernst Haeckel Haus, Institut für Geschichte der Medizin, Naturwissenschaft und Technik, Friedrich Schiller University, Berggasse 7, 07745 Jena, Germany.

actual configuration of the physical surroundings of the receiving system, but to the internal states of this system (Breidbach, Holthausen, and Jost, 1996).

What is this subjective probability? In neuronal networks, a subjective probability is expressed by the instantaneous activation mode onto which an input is superimposed. What does this mean? Local coupling characteristics determine pathways of activation. These do not result in a sampled activity mode, where the overall value of the system's activities is registered, but elicit bulks of periodic oscillations, whose frequency is determined by local coupling functions that are in turn determined by the structural charateristics of nervous tissues. Thus, the system's topology results in a coupling of local dynamics that determine activation patterns characteristic for certain topologies (Holthausen and Breidbach, 1997). An external signal will be superimposed on the resulting activity landscape. Its effect is not directly correlated with the intensity of the input, but depends critically on the situation of the system in which it fits: either it will match a certain oscillation, or it will not. If it fits in, it will strengthen a certain activation pattern; if it does not, the input will either vanish without any effect, or it will change the range of activations of the system. Consequences of such a behaviour might be a re-shaping of the bassins of attraction and, thus, a new oscillation mode of the system will develop. Looking at the attractor configurations, such a change will be expressed in a shift of relative distances between the centers of activation. The elements of such centers will change relative positions, thereby altering the metrics of the system.

A system that reacts in such a way is self-referential. Its internal metric is provided by relative distance functions of cluster elements (Holthausen, 1998; Holthausen and Breidbach, 1999). This internal metric preselects any input situation that is of relevance to the system: only those changes that influence local relations in such a way that not only internal cluster configurations but relative cluster distances that will be affected, are selected. Thereby, the physical description of significant parameter constellations in such relative distance functions allows us to establish a set of rules that the system is going to follow in response to an input situation.

It is argued that one of the basic conceptual schemes of cognitive neuroscience, that of associative psychology (Breidbach, 1997a), can be reformulated in physical terms. The former concept as outlined by James Mill (1869) and later adopted by neuroscientists like Sigmund Exner (1894) and Donald O. Hebb (1949), gave only a general idea about a putative mechanism of brain behaviour. Nevertheless, the idea of Hebb to model the brain using the concept of a neuronal net proves highly successful for cognitive neuroscience: it forms the basis for a neuroscientific interpretation of the actions of neuronal networks. The problem, however, is that, so far, neuroscientific concepts that aim at an analytical interpretation of brain behaviour and the concept of an associative psychology succeed only in a demonstration of structural analogies in their descriptions. The two levels on which those two sciences interpret brain behaviour are not directly interchangeable (Breidbach, 1997b). To correlate these two levels one would have to formulate a theory of the mind/brain, which up to now — in spite of the discussions in philosophy of mind — is not established.

Here, however, another way to gain access for a correlation of associative psychology and an analytical theory of brain functions is proposed. The idea is to transform the concepts of an associative psychology into a physical formalism that will allow an analytical description of what association really is. By that formalism the connectivity of an associative system will be described in an operational way, outlining how far a certain association can be described in terms of functional connectivities of a parallel distributed system. Such a formalism has to establish the rules of associations as physically defined actions of the elements of a nervous system. Such an approach does not lead to a naive reductionism. Using a physical formalism, it introduces a new language into which the descriptions of both levels, the physiological and the psychological level, can be translated. Thus, such a formalism succeeds in a coupling of the two levels of description. In the fomal language offered by physics, it can be precisely demonstrated how far the two levels of description really are correlated — or where they allow mere analogies.

Consequently, a formal treatment of the idea presented is not a matter of choice but is a necessity: The success of such a formal treatment has to be demonstrated in its details. Here, it is argued that the language that allows us to describe the mechanisms of cognition is physics.

Subjective Content

As has been described elsewhere, a self-referential system preselects relevant activity profiles (Becker, 1996; Bell and Sejnowski, 1995; Linsker, 1997; Nadal and Parga, 1994). In such a situation, only a subset of input signals will be effective in eliciting massive system responses (Abeles, 1991). Inputs affect the system not simply by their absolute intensities, but by their relative strength. The input is superimposed on the internal activation pattern of the system (Vaadia, Haalman, Abeles et al., 1995). Accordingly, objective probabilities (p_i) are insufficient to describe the computational processes within such a self-referential system. The system is characterized by system-specific local coupling characteristics, e.g., the interneuronal connections that are specific for the system. Any stimulation of the system will be propagated within this network and any stimulation of the system will be computed according to the preestablished wiring pattern. At any time, furthermore, the system forms an internal activation mode resulting from the overlay of former activations. Thus, in order to describe the information that the system works with, it is not enough to describe a function by which signal patterns of the external world can be transformed within such a system. Because the actual activation pattern of the system is a function of the activation triggered by the stimuli, the preexistent activation pattern and the local coupling characteristics of neurons, one has to incorporate a variable that describes the characteristics internal to the system. This has been done by introducing subjective probabilities (q_i).

The idea to incorporate such internal system characteristics in theoretical neurobiology dates back to 1972. Classical information theory only allows a comparison of a stimulus acting on the system and the output of the system refers to the input (Shannon and Weaver, 1949). Accordingly, one has to know what the input is like. In a typical situation of pattern recognition in the brain, the brain does not know, however, what it is looking at before the pattern has been recognized. Classical information theory had to be extended. This was done by the introduction of subjective probabilities into information theory (Pfaffelhuber, 1972). But, in fact, the mere incorporation of a new type of variable did not solve the problem. What had to be done was to define an internal measure for the activation modes of a system. Important steps in this direction were published by Palm (1981) and Linsker (1988). Palm (1981, 1982) described a function that allowed an approximation of the internal charcateristics of a neuronal network: a system is characterized by the complete number of activation modes it can perform. Because of internal characteristics (weighting of coupling characteristics of an element, threshold level, etc.), each mode will occur with a probability p_x. The surprise function (a physical characterization of the novelty of a message) measures the deviation of the actual probability of a system's activation mode and the putative p_x by which the system is characterized. However, such system characteristics depend on external evaluation.

The question then is whether it is possible to find a measure for the characterization of system behaviour based on internal characteristics. This would allow a definition of information using only internal system characteristics. Shannon and Weaver (1949) have shown that the information content I_x of an event x with the objective probability p_x is given by $I_x = - \log (p_x)$. Kerridge (1961) and Bongard (1970) demonstrated that a subjective probability can be expressed as the representation of the computational predispositions of the system, by using a measure of the subjective content $I_s = - \log (q_x)$. Thus, the information content of a set of events x_i is the average information per event, the Kerridge–Bongard entropy $H_{KB} = \sum_i p_{x(i)} \log (q_{x(i)})$

The subjective probabilities $q_i = q_{x(i)}$ are the basic variables for the system's hypothesis about an external world. Applied to neuronal network theory, the subjective probabilities q_i are defined as functions that depend on network parameters (e.g., synaptic weigths or threshold values). The adaptation of the system was described by Holthausen (1995) as the maximization of H_{KB} using only internal system characteristics. Thereby, the degree of optimization is measured with reference to the pattern of activation assumed by the system in response to a homogenous external world. When using H_{KB} to characterize local coupling functions, measures for the characteristics of self-organizational features can be described. The topology of the system, thereafter, allows a quantification of I_X. In the model presented by Holthausen und Breidbach (1997), maximizing of local information transfer leads to a topologically ordered map of a neural network, whereas the increase of global information fails to do so. The adaptation of the synaptic weights in a self-organizing network, thus, can be described exclusively on the basis of internal variables. Accordingly, these allow a representation of the external world as it is available for the system. The house beetle in a roof construction is not suspected to possess an intrinsic "expert" system for objective features of its habitat (Breidbach, 1986, 1990a). The beetle just has to behave according to internal representations that allow it to react similarily in response to roof beams, old pine wood or telegraph poles.

As has been shown, neuronal networks can be considered as self-determining systems that constitute their own subjective probability distribution by developing an individual topology that predisposes the internal weighting of inputs that enter the system. The subjective probability distribution is re-adjusted in response to an input signal. Integration of sensory information continually changes the system's response characteristics. The question is whether the system's dynamic thereby follows certain rules and, thus, outlines some kind of an internal logic.

What has been demonstrated so far? The concept of an (external) representation, where the system already has to know what it is going to recognize, is insufficient for an analytical definition of association. The latter concept allows us to establish a model for the external decription of behaviour as it provides an expert system into which all known aspects of neuronal control of animal behaviour can be integrated. If this model proves successful, it demonstrates that every relevant mechanism to understand the neuronal basis of a specific behaviour can be outlined. Thus, the model gives information about the completeness of a neurophysiological description. It will, however, not necessarily represent the actual machinery by which the animal brain is working. The same scheme of action can be performed by various neuronal architectures (Breidbach, 1999): comparative neurobiology has already demonstrated this (Breidbach and Kutsch, 1995; Kutsch and Breidbach, 1994). Accordingly, to understand how the brain works, one has to look inside the system and describe its internal characteristics (Ziemke and Breidbach, 1996). An analytical description of a system's intrinsic behavioural pattern can be presented: such a description specifies rules for any parallel computing system with local connectivities and, thereby, also for the brain. To understand the validity of such an approach, its analytical details are of utmost importance — even in a philosophical discussion.

Local Rules for System Behaviour

Via the introduction of local entropy measures, relative distance functions are implemented into the self-referential system: an input changes the actual local activity distribution. There are two possibilities: either the activity landscape corresponding to a part of the system described is invariant under a certain input, or it is not. If it is invariant, the input is without effect on the system; it is just incorporated into the normal oscillation of activity in the system. If it is not incorporated, however, it will change the local coupling functions: it changes the weight vectors of the elements in its surroundings. Because these, likewise, are coupled to neighbouring areas, an eventual effect is propagated — the relative coupling functions of these elements are changed. By measuring the distance to which such an alteration extends (the partition of the phase space affected by such an activation), the relative value of an input for the system can be calculated.

Yet, such a system is a parallel computing system. What does this mean? A sequential machine, like a Turing automaton, will work out a signal recognition procedure by following a line of decision processes. In a parallel computation device, the situation is different. In each instant, activation is dispersed from various components according to their local coupling characteristics. This may result at the second or third step of processing in a complex superposition of activations from various parts of these computing elements. The relative effect of one single activation on the system's reactibility, thus, has to be described regarding the complexity of the system's activation modes. Accordingly, the relative effect of one input on the local activity distribution is registered, measuring its relative effect on neighbouring areas. The effect of an activation, thus, is expressed by the resulting overall distortion in the sequence of locally coupled activity patterns (Holthausen, 1998).

How is the impact of a certain shift in the activation mode of the system to be interpreted? It is necessary to find a relational measure that allows detection of symmetries in a dynamic constellation of changing activity parameters. This can be obtained when it becomes possible to define a measure that describes how the cluster elements interact with each other. A set of coupled elements is called a cluster. Activity modes that establish a stricter coupling of elements, thus, can be regarded as activations of such clusters. Clusters may overlap in some of their elements, but each cluster is characterized by the fact that there is an activation of all elements in one cluster if one of the cluster's elements is activated.

Each cluster element is characterized by its relation to other elements. The elements constitute a cluster when their correlations are significantly closer to each other than to non-cluster elements within a relevant time period. The variation of the local characteristics, thus, must not to be computed as an absolute variation in the metrics of such a coupling group, but as a relative one.

Any signal implemented into the system might distort the former distance values of neighbouring elements. This, however, must not lead to a complete alteration of relative local distances. The analytical definition for clusters, accordingly, has to be based on a measure of the cluster's relative distance functions. Such a measure must allow the computation of relative affinities throughout portions of massive compression or extension of the phase space of activity patterns of a system.

Learning Rules

In applying the principle of local computation, Holthausen (1998) introduced a new learning rule that allowed the study of avalanche effects in local dynamics. His idea was to compute optimal binding characteristics of local elements as functions of activity couplings. Thus, for each moment of a model run of a locally organized system, the coupling functions can be expressed as a transformation into a vector space. One then can form a distance matrix where elements are defined by their relative positions. During a model run, such a matrix that includes all possible coupling functions may oscillate without changing the relative distributions of cluster elements. Looking at the networks of cluster distributions, segregated and aggregated elements can be demonstrated. Thus, a relative measure of coupling intensities can be established (a) for a description of the instantaneous state of the system, and (b) for the characterization of the system throughout a certain period of time. Thereby, clusters again are not defined by the absolute positions of their elements, but by the relative distances of these elements.

Accordingly, this approach does not refer to an external scale. If in the notation developed so far rules for the interaction of system elements can be analytically described, the principal aim envisaged in the problem exposition is fulfilled. Information in a system will no more be defined in regard to the accuracy of representations of an external stimulus, but in regard to the internal activation of the system. Identification of a signal, thus, does not refer to a fixed, externally mediated parameter configuration but is achieved by the dynamic situation of the system itself.

Within a cluster, a central group of cluster elements can be defined that shows interactions only with other cluster elements. Furthermore, a border group of elements is characterized. These are not exclusively coupled to elements of only one cluster. There can be a tendency of such border elements either to become integrated in or to be separated from a certain cluster. Thereby, even a hierarchy within the set of clusters can be established: those clusters that lose border elements to another cluster can be regarded as being dominated by that cluster et vice versa.

The learning rule by which the behaviour of such a system can be described is defined by the frequency of adaptations (Khaikine and Holthausen, 1999). The latter is determined by an objective probability vector: in Holthausen's (1998) approach this represents only the coupling between a given input distribution and a mapping function — only the relative distances of two maps are computed. The resulting function is self-referential, as it depicts only characters intrinsic to the system.

The Holthausen algorithm presents not only the introduction of order characteristics into a dynamical system, but is successful in establishing rules that allow the identification of classes in the activity patterns of neuronal networks. The performance of the system is implemented in the dynamics of its local topologies: the relation of single cluster elements to each other. A certain cluster, thus, can be defined in such a way that a relative distance matrix is established that gives weighting factors for the correlation of the clusters to each other. Thus, it is possible to express degrees of distance between certain clusters. By a simple Boolean operation in the vector space characterizing the distance functions of every cluster element, clusters that overlap (to a certain degree) in the activation of other elements can be identified. By a description of the relative distances of the elements, the relations of each cluster to all other clusters are expressed. The rules for the interactions of the clusters, thus, are defined by the microdynamics of their elements. These elements are attractors, defined by the local dynamics of the system. Accordingly, the gross characteristics of the system's behaviour can be traced down to the level of the system's local coupling dynamics, and described in the language of physics.

Rules for System Intrinsic Interactions

The relation of two elements is described by a distance function. The elements are part of a certain cluster or they are not. When the formation of such a cluster is changed, the position of an element with regard to such a cluster can alter. If the history of the various cluster-configurations in which two elements were found is registered, the dynamics of the cluster relations can be analysed. An element can be part of two clusters a,b showing a changing tendency of cluster attribution thoughout a model run. That changing attribution can be measured. This analysis can be done for each element of cluster *a* and cluster *b*, resulting in a matrix detailing not only the relations between elements of cluster *a* and *b* but also the overall distances of both clusters. Furthermore, in an analysis of the dynamical behaviour of clusters, a simple analysis of the growth or shrinkage of the numbers of elements gives information about the relative intensity of certain cluster couplings. The simple rules that can be found by an analysis of the relations of certain cluster elements do not allow only the identification of similar elements but allow us even to establish degrees of similarities. Such an analysis will provide to us the framework for a logic for the behaviour of neuronal networks. To understand this properly, one has to understand what a cluster element really is.

We do not look at the coupling functions expressed in the structural organization of knots and grids in a neuronal network. The analysis performed works on a more abstract level. A cluster element reflects underlying knot activations, but it is not identical with them (Holthausen and Breidbach, 1997). An element, as it is presented here, represents the underlying coupling and activation modes of a network area (eigenvectors of activity patterns). The activity pattern of the system may be expressed by varying local coupling functions of neurons and by varying time characteristics of the actual binding functions. All these varying activation patterns elicit a local dynamic that can be studied in an analysis of attractor dynamics (Holthausen and Breidbach, 1999). The topology of the dynamical system is, thus, not the actual network activity pattern, but the series of attractors that were elicited by these activation modes. In consequence, relations between two cluster elements do not correspond directly to transformations of weighting functions for synaptic intensities and threshold values. The relative distances measured in our approach describe the attractor dynamics. The cluster constitutes that attractor. Accordingly, the cluster will change its actual metric, if it is compressed. The cluster will not lose its relative characteristics, however, if it can be described as isomorphous (to itself) throughout its eventual transformations.

This sketch demonstrates why it is advantageous to work with such a complicated procedure that avoids absolute scaling of activation or coupling characteristics of or within system activation patterns. Using such clusters, even deformed attractors can be identified. Thus, such an analysis allows the detection of similarities and dissimilarities even in compressed or extended partitions of phase spaces. Consequently, the formalism developed does not depend on fixed parameters but is invariant under changes in local parameter constellations — the tolerated degrees of such variabilities are defined. Local fluctuations that shatter the constellation of cluster elements are identified as such a limitation. These do not depend on fixed parameters, as the latter vary with respect to the actual overall distribution of activations within the system. The criteria gained for isomorphous structures, thus, are not simple. Isomorphy, accordingly, is a relationally defined constellation — isomorphs are such constellations that can be implemented in an identical cluster. The cluster is defined by the relative positioning of its elements. The question, thereafter, is, up to what degree are two transformable modes in the development of one cluster no longer seen to reflect an isomorphous constellation. This situation is given when the transformation can only be followed in one direction, that is, in a bifurcation (Pasemann, 1995).

Accordingly, the elements of the logic of relations presented are not strictly defined: only a relational characterization is given. The picture depicted here describes a network of interactions resulting in an interwoven grid of vector functions characterizing different activity states of local areas in the system. It has been shown that the resulting topological characteristics of system activation patterns suffice to characterize the internal dynamics of the system (Holthausen and Breidbach, 1997).

Clusters and Internal Logic

A cluster C of elements *e* is defined throughout a time period. A logic of relations should allow the characterization of dynamical changes over various time periods. Such a logic is based on considerations about the interdependencies of elements. Even the definition of identity has to be formed in a relational manner. Thereby, the problem is that elements show continual fluctuations of their absolute positions in the phase space. An absolute scaling would not allow us to define self-identities of such dynamic elements. How then, can the trace of such an element in phase space be followed? To do that, a relative measure, by which the position of an element can be defined, has to be introduced. Its definition has to refer to relational characteristics: identity of a set of elements over a time period is defined in reference to other elements.

Hereby, a relative coordinate system is established whose dimensions vary according to the state of activation of the complete system. Consequently, the traces of single elements can be followed throughout the history of the system (Holthausen, 1998).

An alteration in the absolute scaling of the relative positions of various elements in such a system — as an effect of an incorporation of a new element in the system — does not distort the principal topology of the system: the system is defined by the relative positions of its elements to each other. The introduction of a new element changes the system's intrinsic relations and, thus, affects — more or less — all elements. If the change induced by such an introduction is a considerable one, the system is affected in each part; consequently, it will not fade but follow a common trend. Such trends can be evaluated by looking at various physical characteristics of the system (Holthausen and Breidbach, 1999).

James Mill (1869) had outlined how in a network of interwoven oscillators within the brain sensations overlay each other and, thus, modify the topology of the signal processing system. His approach is suggestive, especially, when it is transformed into a functional morphological concept in neuroscience where the brain is described as a network of interwoven neurons. Such an idea of neuroanatomy was already established in the last decades of the nineteenth century, where — consequently — Mill's concept found tremendous respect (Breidbach, 1997a). Nevertheless, in those years the rigour of a physical theory which presented a mechanism by which associative features really could be understood was lacking.

It must be demonstrated how a new input can be integrated in a system. Thereby, it should be made clear that a new input will not only add one particular quality into the signal detecting system, but that it will optimize the general resolution capacity of the system.

Let us describe the system behaviour in terms of relational characteristics. By the implementation of a new element into the system, the scaling of relative distance functions in the system is magnified: there will be one further element in each matrix by which the functional characteristics of the system are described. Accordingly, the accuracy by which the relative distances of the various elements in the matrix are expressed, is improved. Thus, when a new element is introduced, the relations of the elements to each other will be automatically re-evaluated regarding the new matrix functions. Consequently, by introducing more and more elements in such a relationalistically characterized system, the system will gradually optimize its accuracy in separating more closely related clusters. Mathematically, this can be described by formulae for the capacity and complexity of the system (Jost, 1998).

The rules for the combinations of clusters in a system can be described. These rules allow the outlining of a logic of system activation modes. As these can be traced down to the level of single elements (as forming part of the cluster), a neuro-logic would be established, at least in principle. Thus, a framework for a physical theory of cognition would be secured. Below, formulae are sketched that might establish such a framework. It is reasonable to follow the technical details, as these outline whether and how far such a framework can be established.

Internal Logic

An activation x in the system is reflected in a complex relation, by which the event x is transferred throughout the network of related elements. After a time period the resulting activity pattern of the system cannot easily be traced back to the situation at the beginning of the period. The matrix of distances has changed completely. The original parameter constellation characterizing *x* at the beginning of signal processing is lost (Holthausen, 1998) — a static definition of identity is useless in such a dynamical situation. Identity is found within an element's activation when the same relations between the elements of a certain cluster are expessed.

In the dynamical system, each element continually integrates new activations. Thereby, there is a kind of continuous check of relational characteristics within each of the elements. In the central region of a cluster the hierarchy of relations within the cluster elements is conserved during the time of computation, whereas border elements tend to become temporarily part of another cluster.

The attachment of border elements to cluster C during a time interval depends on the relative probability of coupling rates. If the element has a significantly higher binding rate to the elements or part of the elements of one cluster, it is part of this cluster. The less significant such coupling characteristics are, the less obvious is the attachment of such an element to a particular cluster.

Two clusters interact when there are some elements shared by both. If the number of shared elements increases, two clusters become more similar. Two clusters which do not share central elements are different. Different clusters which share all border elements are similar. Similarity decreases with the decrease in the number of shared border elements.

By an analysis of such rules of system dynamics, an operationalistic approach to logic becomes feasible. In a first step, only an elementary logic including the junctors "and," "or," "implication," and the all-quantor (*x* is true for all elements) is aimed at. As Frege (1966) has outlined, logical operations can be reduced to these basic procedures. It is possible to describe such a logical scheme for relations between clusters. Elements of the logical operations, thus, are relations between clusters and not only states of certain clusters.

If cluster *A* and cluster *B* are different and an activation of cluster *A* is followed by an activation of cluster *B* and, likewise, an activation of cluster *B* is followed by an activation of cluster *A*, the situation can be regarded as an "and" function. If cluster *A* and cluster *B* are different but share border elements, and one of these clusters is activated — but this activation is never followed by activation of the second cluster — this can be regarded as an exclusive "or" function.

If cluster *A* and cluster *B* are different and an activation of cluster *A* is followed by an activation of cluster *B*, but an activation of *B* does not activate cluster *A*, this can be regarded as corresponding to a logical implication.

If within a number of clusters A_N a certain cluster is activated and such an activation is followed by an activation of all clusters A_N, this can be regarded as corresponding to an all-quantor operation.

In this way, a calculus is established that represents a kind of logic in parallel computing systems. Similarity, difference, isomorphy, hierarchy and implication became definable in terms of network activity modes. Thus, a predicate logic becomes feasible. In addition, Holthausen and Breidbach (1997) demonstrated that a system's topology defines subjective contents: rules describe the basic semantics of the elements that constitute a system. What has to be proven, however, is that this approach is sufficient to represent categories by which internal data are ordered in such a way that they form a semantic system. Of course, the categorization of a stimulus representation is the effect of a system's intrinsic mechanisms (Breidbach, 1997a). Also demonstrated is the close correspondence of the behaviour of such a system to a basic outline of associative psychology such as that described by Mill (1869). What is lacking is a complete mathematical description of such categorization algorithms. Model runs, however, demonstate that the described framework of a neuro-logic "works" (Holthausen, 1998): using such algorithms, it is possible to implement self-organizing cluster functions by which a system forms its individual topology (Holthausen and Breidbach, 1997).

Internal Logic and Neurosemantics

The rules presented here are physical characterizations of the behaviour of parallel computing systems. Following the concept of internal representations, within the analysis of network dynamics, a framework of basic logical functions is obtained. The physics applied reflects the principal characteristics of neuronal network organizations: their local characteristics and the highly stereotyped morphology of neurons. A physical model that reflects these central attributes of natural neuronal systems is sufficient to establish a kind of system intrinsic logic.

There is one striking parallel in the analysis of the physical model here presented and in the psychological concepts that dominated nineteenth century brain physiology. The historical development of the concept of an associative mind that was generally agreed upon at the time of Sherrington and his followers cannot be described here in detail (see Breidbach, 1997a). But what is demonstrated is that Mill's system of associative psychology is analogous to the descriptions of the behaviour of a parallel distribution system with local characteristics. According to Mill, the first effects of an impression in the brain are not a proper representation of the physical world. The effects are described as being the outcome of an interference of inputs with the internal activity mode of the system (Breidbach, 1996). Mill, and later the physiologist Sigmund Exner (1894), saw signal identification and association as the result of endogeneous brain activation modes. Signal inputs elicit activation in certain neuronal pathways; these are superimposed on internal oscillations and, thus, cause more complex reactions in brain tissue. The result is a complex coactivation of signal pathways established by former impressions. If such coactivations can be performed in different modes, for example, if an activation corresponding to red is elicited, likewise, by a cherry and a rose, the coactivated attractor is likely to be regarded as a general attribute (in this case, the attribute red) of a set of input situations. Thus, Mill presented a coherent picture of a possible physiological background of cognitive actions like identification, association and even memory.

The relationalistic definition of the system, employing internal representation, presents the description of associations in an analytical way. Here, entities like "cherry" and "apple" are regarded as a cluster of elements. In the technical representation of such an associative system, as has been described in the present paper, elements correspond to attributes putatively activated by various clusters. The relative distances of these elements correspond to the divergence of various clusters. The introduction of a new element into such a relative mapping configuration may rearrange the cluster distribution and may even extend the general resolution capacity of the system. Rules are outlined by which regularities in the cluster interaction are to be formalized. These rules show that an elementary logic can be implemented in the activation modes of the system. By outlining the rules of cluster interaction in such a system its semantics are described.

A stimulation of elements in the parallel computing system results in a temporary shift of its activation modes. These activation patterns are processed in the interneuronal connections. If an activation shift is stabilized over a longer time period, it will establish new modes of local interactions. These dynamics are described as shifts in the activity distribution of the system shown as oscillation patterns in a phase room. A single oscillation pattern corresponds to a single attractor. Since the elements of the system's activation modes can be physically defined as distinct attractors, physically the system is described by its attractor configuration. A complete understanding of what is going on on the system level, however, has to describe transient dynamics, that is, the microdynamics underlying the attractor characteristics. James Mill had envisaged such a physical description and was able to develop his idea of coactivations as a principal scheme for an understanding of the physiological basis of associations — thereby he speculated about the origins of the categories we use in our verbal analysis of the world as well as ourselves (Kurthen, 1992). Mill portrayed a relationist's view of the mental representation of the world. The mechanisms he proclaimed as being effective *were* the associations.

Caution has to be applied, however, when trying to parallel a modern account against Mill's speculative ideas. Thus, the present account starts with the field of physics which demonstrates basic characteristics of parallel computation. The formulation of rules for activation modes characterizes particular intrinsic attributes of a system (Holthausen, 1998; Holthausen and Breidbach, 1999; Khaikine and Holthausen, 1999). The crucial point is seen in the topology of the system (Holthausen and Breidbach, 1997). By a dynamical interpretation of the elements that constitute the topology of the system, a new purely instrinsic definition of information is given (Holthausen, 1998). The description of a system's intrinsic computational characteristics is sufficient to reflect at least the framework of a logical calculus. Such characteristics are not necessarily neuronal. Physics allowed the demonstration of principal qualities of systems possessing such characteristics. The nervous systems is just one of these.

References

Abeles, M. (1991). *Corticonics. Neural circuits of the cerebral cortex.* Cambridge and New York: Cambridge University Press.

Ashby, W.R. (1962). What is mind? Objective and subjective aspects in cybernetics. In M.J. Scher (Ed.), *Theories of the mind* (pp. 305–313). New York: Free Press.

Averof, M., and Akam, M. (1995). Hox genes and the diversification of insect and crustacean body plans. *Nature, 388*, 682–686.

Averof, M., Dawes, R., and Ferrier, D. (1996). Diversification of arthropod HOX genes as a paradigm for the evolution of gene functions. *Seminars in Cellular and Developmental Biology, 7*, 539–551.

Bastiani, M.J., Doe, C.Q., Helfand, S.L., and Goodman, C.S. (1985). Neuronal specificity and growth cone guidance in grasshopper and Drosophila embryos. *Trends in Neurosciences, 4*, 257–266.

Becker, S. (1996). Mutual information maximization: Models of cortical self-organization. *Network: Computation in Neural Systems, 7*, 7–31.

Bell, A.J., and Sejnowski, T.J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*, 1129–1159.

Bongard, M. (1970). *Pattern recognition.* New York: Spartan Books.

Braitenberg, V. (1973). *On the texture of brains.* New York: Springer Verlag.

Braitenberg, V. (1984). *Vehicles. Experiments in synthetic psychology.* Cambridge, Massechusetts: MIT Press.

Braitenberg, V., and Schüz, A. (1991). *Anatomy of the cortex.* Berlin: Springer Verlag.

Breidbach, O. (1986). Studies on the stridulation of *Hylotrupes bajulus* (L.) (*Cerambycidae, Coleoptera*): Communication through support vibration. *Behavioural Processes, 12*, 169–186.

Breidbach, O. (1990a). Zur Struktur des Aggressionsverhalten des Cerambyciden *Hylotrupes bajulus* L. (*Col. Cerambycidae*). *Deutsche entomologische Zeitschrift, Neue Folge, 37*, 23–30.

Breidbach, O. (1990b). Constant topological organization of the coleopteran metamorphosing nervous system: Analysis of persistent elements in the nervous system of *Tenebrio molitor. Journal of Neurobiology, 21*, 990–1001.

Breidbach, O. (1990c). Reorganization of persistent motoneurons in a metamorphosing insect (*Tenebrio molitor* L., *Coleoptera*). *Journal of Comparative Neurology, 302*, 173–196.

Breidbach, O. (1996). Vernetzungen und Verortungen. Bemerkungen zur Geschichte des Konzeptes neuronaler Repräsentation. In A. Ziemke and O. Breidbach (Eds.), *Repräsentationismus — Was sonst?* (pp. 35–62). Braunschweig: Vieweg.

Breidbach, O. (1997a). *Die Materialisierung des Ichs. Zur Geschichte der Hirnforschung im 19. und 20. Jahrhundert.* Frankfurt: Suhrkamp.

Breidbach, O. (1997b). Denken in Neuronalen Netzen. In K.P. Dencker (Ed.), *Labile Ordnungen. Interface III* (pp. 40–53). Hamburg: H. v. Bredow.

Breidbach, O. (1999). Comparing minds — A comment on beetles' intelligence. *Theory in Biosciences, 118,* 54–65.

Breidbach, O., Dircksen, H., and Wegerhoff, R. (1995). Common general morphological pattern of peptidergic neurons in the arachnid brain: Crustacean cardioactive peptide-immunoreactive neurons in the protocerebrum of seven arachnid species. *Cell and Tissue Research, 279,* 183–197.

Breidbach, O., Holthausen, K., and Jost, J. (1996). Interne Repräsentationen — Über die "Welt"generierungseigenschaften des Nervengewebes. Prolegomena zu einer Neurosemantik. In A. Ziemke and O. Breidbach (Eds.), *Repräsentationismus — Was sonst?* (pp. 177–196). Braunschweig: Vieweg.

Breidbach, O., and Kutsch, W. (1995). *The nervous system of invertebrates: An evolutionary and comparative approach.* Basel: Birkhäuser.

Exner, S. (1894). *Entwurf zu einer physiologischen Erklärung der psychischen Erscheinungen.* Leipzig: F. Deuticke.

Frege, G. (1966). *Logische Untersuchungen.* Göttingen: Vandenhoeck und Ruprecht.

Gaze, R.M. (1970). *The formation of nerve connections.* London: Academic Press.

Goodman, C.S. (1982). Embryonic development of identified neurons in the grasshopper. In N.C. Spitzer (Ed.), *Neuronal development* (pp. 171–212). New York: Plenum Press.

Hebb, D. O. (1949). *The organization of behaviour: A neuropsychological theory.* New York: Wiley.

Holthausen, K. (1995). *Neuronale Netzwerke und Informationstheorie.* Münster: Ph.D. Thesis, University of Münster.

Holthausen, K. (1998). Evolution of internal representations generated by unsupervised self-referential networks. *Theory in Biosciences, 117,* 18–31.

Holthausen, K., and Breidbach, O. (1997). Self-organized feature maps and information theory. *Network: Computation in Neural Systems, 8,* 215–227.

Holthausen, K., and Breidbach, O. (1999). Analytical description of the evolution of neural networks: Learning rules and complexity. *Biological Cybernetics, 8,* 215–227.

Hoyle, G. (1975). Identified neurons and the future of neuroethology. *Journal of Experimental Zoology, 194,* 51–74.

Jost, J. (1998). On the notion of complexity. *Theory in Biosciences, 117,* 161–172.

Jost, J., Holthausen, K., and Breidbach, O. (1997). On the mathematical foundation of a theory of neural representation. *Theory in Biosciences, 116,* 125–139.

Kerridge, D.F. (1961). Inaccuracy and inference. *Journal of the Royal Statistitical Society, B 23,* 184–194.

Khaikine, M., and Holthausen, K. (1999). A general probability estimation approach for neural computation. *Neural Computation, 12,* 457–474

Kurthen, M. (1992). *Neurosemantik.* Stuttgart: Enke.

Kutsch, W., and Breidbach, O. (1994). Homologous structures in the nervous systems of arthropoda. *Advances in Insect Physiology, 24,* 1–113.

Kutsch, W., and Heckmann, R. (1995). Homologous structures, exemplified by motoneurons of Mandibulata. In O. Breidbach and W. Kutsch (Eds.), *The nervous systems of invertebrates: An evolutionary and comparative approach* (pp. 221–248). Basel: Birkhäuser.

Kutsch, W., Urbach, R., and Breidbach, O. (1993). Comparison of motor patterns in larval and adult stage of a beetle, *Zophobas morio. Journal of Experimental Zoology, 267,* 389–403.

Linsker, R. (1988). Self-organization in a perceptual network. *Institute of Electronic and Electric Engineer (Computer), March,* 105–117.

Linsker, R. (1997). A local learning rule that enables information maximization for arbitrary input distribution. *Neural Computation,* 9, 1661–1665.

Macnamara, J., and Reyer, G.E. (Eds.). (1994). *The logical foundations of cognition.* New York: Academic Press.

Mill, J. (1869). *Analysis of the phenomena of the human mind.* London: Longman's.

Nadal, J.-P., and Parga, N. (1994). Non-linear neurons in the low noise limit: A factorial code maximizes information transfer. *Network: Computation in Neural Systems,* 5, 565–581.

Palm, G. (1981). Evidence, information, and surprise. *Biological Cybernetics*, *42*, 57–68.

Palm, G. (1982). *Neural assemblies*. Berlin: Springer.

Pasemann, (1995). Neuromodules: A dynamical system approach to brain modelling. In H.J. Herrmann, D.E. Wolf, and E. Pöppel (Eds.), *Supercomputing in brain research: From tomography to neural networks* (pp. 331–348). Singapore: World Scientific.

Penrose, R. (1989). *The emperor's new mind*. Oxford: University Press.

Pfaffelhuber, E. (1972). Learning and information theory. *International Journal of Neuroscience*, *3*, 83–88.

Rusch, G., Schmidt, S.J., and Breidbach, O. (Eds.). (1996). *Interne Repräsentationen*. Frankfurt: Suhrkamp.

Shannon, C.E., and Weaver, W. (1949). *The mathematical theory of communication*. Champaign–Urbana: University of Illinois Press.

Shubin, N., Tabin, C., and Caroll, S. (1997). Fossils, genes and the evolution of animal limbs. *Nature*, *361*, 490–492.

Vaadia, E., Haalman, I., Abeles, M., Bergman, H., Prut, Y., Slovin, H. and Aertsen, A. (1995). Dynamics of neuronal interactions in the monkey cortex in relation to behavioural events. *Nature*, *373*, 515–518.

Ziemke, A., and Breidbach, O. (Eds.). (1996). *Repräsentationismus — Was sonst?* Braunschweig: Vieweg.